Analyse de la variance à un facteur (ANOVA) : comparaison de k moyennes sur des groupes indépendants

Exercice 14

Dans un établissement scolaire, on a réparti les élèves en trois classes de troisième ; les notes ci-dessous sont celles obtenues par les élèves en mathématiques au Brevet des Collèges. Peut-on dire que ces trois classes sont équivalentes ? Si oui, quelles seraient les caractéristiques de la population résultant de la fusion des trois groupes ?

G1	G2	G3
14	8	7
15	18	8
20	3	11
7	12	11
8	15	20
13	8	14
10	7	13
1	11	13
12	8	10
16	14	12
17	14	12
17	9	13
11	9	12
6	9	14
16	10	8

G1	G2	G3
8	14	13
10	15	12
11	14	8
11	13	8
7	10	11
10	12	15
11	10	8
12	12	14
11	12	16
8	11	13
	10	12
	10	15
	10	
	12	

Vérifier l'exactitude des tableaux ci-dessous et conclure.

	G1	G2	G3	Totaux	
n_j	25	29	27	81	
T_j	282	320	323	925	10563,27
Σx_{ij}^2	3600	3782	4091	11473	
T_j^2/n_j	3180,96	3531,03	3864,04	10576,03	
Inter	12,76				
Total	909,73				

Sources de variations	Sommes des carrés	DDL	Carrés moyens	F
Inter	12,76	2	6,38	0,55
Intra	896,97	78	11,50	
Total	909,73	80		

Réponses : Au seuil de 5%, $F_{crit}(2,78) = 3.1$. La différence entre les groupes n'est donc pas significative. De plus, l'obtention d'un F_{obs} inférieur à 1 semblerait indiquer (sans pour autant le montrer) que les classes n'ont pas été constituées au hasard, mais qu'elles ont, au contraire, été rendues artificiellement homogènes : on a composé les trois classes de façon qu'elles soient de niveau équivalent.

Exercice 15

Lors d'une expérience pédagogique, on s'intéresse à l'effet comparé de deux pédagogies des mathématiques chez deux groupes de 10 sujets :

- pédagogie traditionnelle (p_1)
- pédagogie moderne (p_2)

On note la performance à une épreuve de combinatoire.

p_1 traditionnelle	
s1	5.0
s2	4.0
s3	1.5
s4	6.0
s5	3.0
s6	3.5
s7	3.0
s8	2.5
s9	1.5
s10	2.5

p_2 moderne	
s11	4.0
s12	5.5
s13	4.5
s14	6.5
s15	4.5
s16	5.5
s17	1.0
s18	2.0
s19	4.5
s20	4.5

1) Vérifier que les paramètres des deux échantillons sont donnés par :

	p_1	p_2
Moyenne	3.250	4.250
Ecart-type	1.365	1.553
Variance	1.863	2.413
Ecart-type corrigé	1.439	1.637
Variance corrigée	2.069	2.681

2) Ces données expérimentales permettent-elles d'affirmer que la pédagogie a un effet sur les résultats à l'épreuve de combinatoire ?

a) Comparer les moyennes des deux groupes à l'aide d'une analyse de variance.

b) Comparer les résultats avec ceux obtenus au premier semestre, à l'aide de la statistique T.

Réponses :

Les calculs intermédaires sont résumés dans le tableau suivant :

	Péda1	*Péda2*	*Totaux*	
n_j	*10*	*10*	*20*	
T_j	*32.5*	*42.5*	*75*	*281.25*
Σx_{ij}^2	*124.25*	*204.75*	*329*	
T_j^2/n_j	*105.625*	*180.625*	*286.25*	
Inter	*5.00*			
Total	*47.75*			

Le tableau d'analyse de variance est donc :

Sources de variation	*Sommes des carrés*	*DDL*	*Carrés moyens*	*F*
Inter	*5,0*	*1*	*5,0*	*2,11*
Intra	*42,75*	*18*	*2,375*	
Total	*47,75*	*19*		

Au seuil de 5%, $F_{crit}(1,18) = 4.41$. Hypothèse H_1 rejetée.
Comparaison possible avec l'exercice vu au premier semestre : $t^2_{obs} = (-1.45)^2 = 2.10$, c'est-à-dire la valeur de F.

Enoncé 16 *Données Bransfor*
On reprend une expérience de Bransford et al. (1972), dans laquelle on demande à des sujets d'écouter le texte suivant :

> "Si les ballons éclatent, le son ne portera pas puisque tout sera bien trop loin du bon étage. Une fenêtre fermée empêchera également le son de porter, surtout depuis que les immeubles récents sont correctement isolés. Comme l'essentiel de l'opération dépend d'une arrivée correcte d'électricité, un fil cassé causerait bien des problèmes. Evidemment, le type peut hurler. Mais la voix humaine n'est pas assez puissante pour porter bien loin. Un problème supplémentaire serait qu'une corde casse sur l'instrument. Alors il serait impossible d'accompagner le message. C'est clair que la meilleure situation impliquerait la plus petite distance. Alors, il y aurait bien moins de problèmes potentiels. Avec un contact en face à face, un bien petit nombre de choses pourrait gêner."

Le but visé par Bransford *et al.* est de montrer l'importance du contexte dans la compréhension et la mémorisation d'un texte. Pour ce faire, ils utilisent quatre groupes expérimentaux :

1. Un groupe "sans contexte" entend simplement le texte.
2. Le groupe "avec contexte avant" regarde une figure suggérant un contexte approprié pendant qu'il entend le texte.
3. Le groupe "avec contexte après" entend le texte puis regarde la figure précédente.
4. Le groupe "avec contexte partiel" regarde une figure suggérant un contexte inapproprié pendant qu'il entend le texte.

A proprement parler cette étude comprend un groupe expérimental (le groupe 2 : contexte pendant) et trois groupes contrôles (les groupes 1, 3 et 4). Les groupes contrôles doivent permettre d'éliminer des explications concurrentes (en particulier, effet facilitateur sur la mémoire de l'imagerie, de l'aspect concret du matériel, etc.). L'expérimentateur s'attend, donc, à observer une performance pour le groupe 2 supérieure aux trois autres groupes.
Il choisit de mesurer le comportement des sujets par deux Variables Dépendantes : une note de compréhension donnée par les sujets (de 0 à 7, avec 0 indiquant l'incompréhension totale), et le nombre d'idées correctement rappelées (Bransford découpe le texte en 14 idées, essayez de les retrouver !). Quoique cette dernière Variable Dépendante soulève de délicats problèmes de codage (e.g., à partir de quel moment une idée est présente...), elle reflète clairement l'intérêt des auteurs de cette expérimentation.
Dans cette expérience, on utilise vingt sujets répartis en quatre groupes. Les résultats, pour la Variable Dépendante "nombre d'idées rappelées" (maximum 14) se trouvent ci-dessous (mais avant, faites ce que doit faire un bon expérimentateur : prenez une feuille et détaillez les cinq premières étapes du test statistique avant de partir à la pêche aux résultats) :

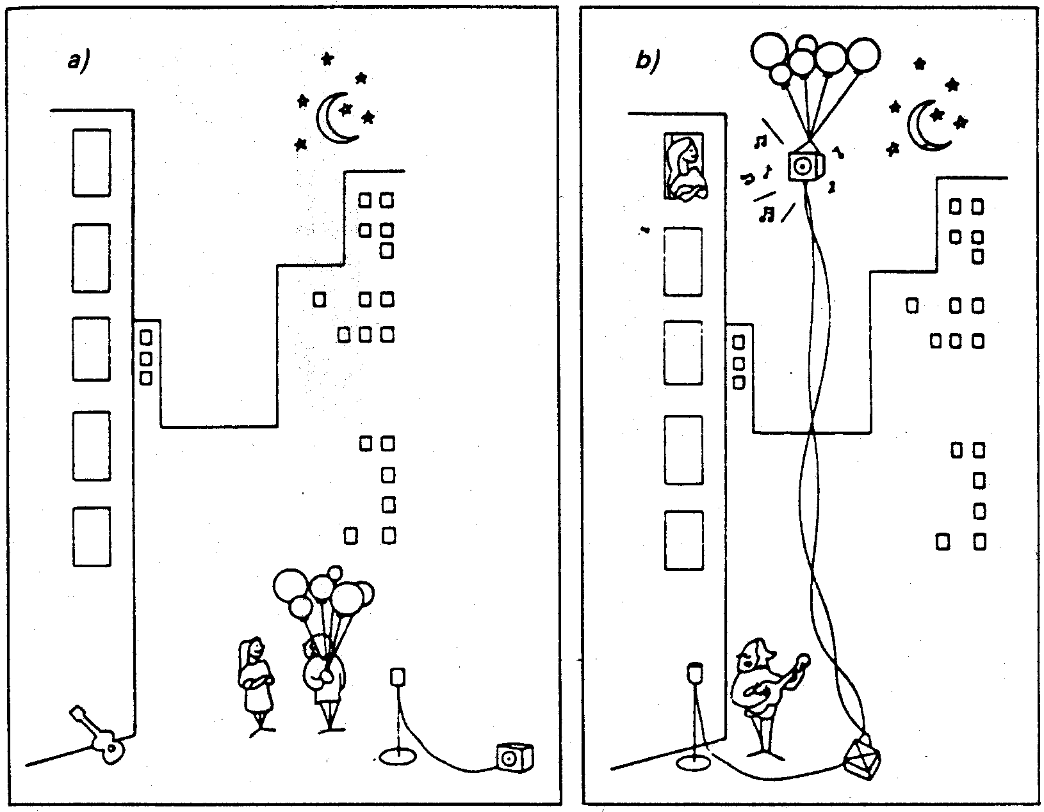

FIG. 1 – Contexte inapproprié (a) et approprié (b) pour l'expérience de Bransford

Résultats de l'expérience				
	G.1	G.2	G.3	G.4
	3	5	2	5
	3	9	4	4
	2	8	5	3
	4	4	4	5
	3	9	1	4
$T_{.j}$	15	35	16	21
n_j	5	5	5	5
$\frac{T_{.j}}{n_j}$	3	7	3.2	4.2
Σx_{ij}^2	47	267	62	91

Justifiez les calculs et le tableau d'ANOVA suivants :
Table d'ANOVA :

Source	ddl	SC	CM	F_{cal}	$Pr(F > F_{cal})$
$\mathcal{A}$	3	50.95	16.98	7.22 **	.00288
$\mathcal{S}(\mathcal{A})$	16	37.60	2.35		
Total	19	88.55			

Si on utilise la procédure des valeurs critiques :
** $F_{critique} = 5.29$, au seuil $\alpha = .01$; $F_{cal} > F_{critique}$. On rejette H_0.
Les cinq étapes du test sont évidemment :

1. *Formulation des hypothèses statistiques H_0 et H_1. Ici :*
   H_0 : dans les 4 conditions, les moyennes dans la population parente sont égales
   H_1 : les 4 moyennes ne sont pas toutes égales.
2. *Choix du test : ici, une analyse de variance à un facteur. Statistique : F.*
3. *Distribution de la statistique de test : ici, le F de Fisher Snedecor avec $ddl_1 = 3$ (nombre de groupes – 1) et $ddl_2 = 16$ (nombre d'observations – nombre de groupes).*
4. *Seuil de signification choisi : ici, $\alpha = 1\%$.*
5. *Règle de décision : détermination des zones d'acceptation et de rejet de H_0. Ici, :*
   – *Si $F_{cal} \leq 5.29$, on accepte H_0 (égalité des moyennes)*
   – *Si $F_{cal} > 5.29$, on refuse H_0 et on accepte H_1.*

L'étude pourrait être poursuivie à l'aide de la méthode des contrastes orthogonaux (que nous ne détaillerons pas).
La première étape consiste opposer le groupe 2 aux trois autres groupes en testant l'hypothèse nulle : $3\mu_2 = \mu_1 + \mu_3 + \mu_4$. On calcule : $L_1 = 3\overline{x_2} - \overline{x_1} - \overline{x_3} - \overline{x_4} = 10.6$; $\sum a_j^2 = 3^2 + (-1)^2 + (-1)^2 + (-1)^2 = 12$; $SC_{contraste1} = \dfrac{nL^2}{\sum a_j^2} = 46.81$
Dans la formule précédente, n est le nombre d'observations par groupe. Ici, $n = 5$. Le F de Fisher associé à ce contraste est obtenu en divisant $SC_{contraste1}$ par le carré moyen résiduel 2.35 ; il vaut 19.92. Les degrés de liberté sont 1 et 16. Le résultat est donc significatif d'un comportement du groupe 2 différent de celui des autres groupes.
La méthode peut être poursuivie en opposant le groupe 4 aux groupes 1 et 3 (coefficients appliqués aux quatre moyennes : 1, 0, 1, -2) puis en opposant les groupes 1 et 3 (coefficients appliqués : 1, 0, -1, 0).
Pourquoi s'agit-il de contrastes orthogonaux *?*
Réponse : Les "vecteurs" associés aux coefficients des trois contrastes, à savoir $V_1 = (-1, 3, -1, -1)$, $V_2 = (1, 0, 1, -2)$, $V_3 = (1, 0, -1, 0)$ sont deux à deux orthogonaux (par exemple, $V_1 \cdot V_2 = -1 \times 1 + 3 \times 0 + (-1) \times 1 + (-1) \times (-2) = 0$), ce qui garantit l'indépendance des résultats des trois tests.
Une autre grandeur intéressante est le coefficient (souvent noté η^2) d'estimation de l'intensité de l'effet de la variable indépendante. Dans le cas d'une analyse de variance à un facteur, il est défini par :

$$\eta^2 = \frac{SC_{inter}}{SC_{total}}$$

Il vaut donc ici : $\eta^2 = 0.58 = 58\%$.
Signification : 58% de la variance de la Variable Dépendante est expliquée par la Variable Indépendante (les différentes conditions expérimentales).
η^2 est aussi le carré d'un coefficient de corrélation. η peut en effet être obtenu comme coefficient de la corrélation entre l'ensemble des données observées d'une part, et la série de données obtenue en remplaçant chaque observation par la moyenne de son groupe d'autre part. Sur notre exemple, soit U la série des données observées et V la série des données du modèle ainsi obtenu.

u_i	3	3	2	4	3	5	9	8	4	9	2	4	5	4	1	5	4	3	5	4
v_i	3	3	3	3	3	7	7	7	7	7	3.2	3.2	3.2	3.2	3.2	4.2	4.2	4.2	4.2	4.2

On obtient : $r(U,V) = 0.7585$ *et* $r^2(U,V) = 0.575$.

Enoncé 17 *Données Loftus*
Elisabeth Loftus (Loftus et Palmer 1974) —– dans une série d'expérimentations sur le thème du témoignage —– désire mettre en évidence l'influence de la tournure d'une question sur la réponse de témoins. Pour ce faire, elle montre à ses sujets, un film décrivant un accident de voiture. Elle pose, ensuite, une série de questions aux sujets. Parmi celles-ci se trouve une des cinq versions d'une question relative à la vitesse des véhicules. Voici ces versions :

1) HIT : About how fast were the cars going when they *hit* each other ? (A environ quelle vitesse allaient les voitures quand elles se sont "percutées").

2) SMASH : About how fast were the cars going when they *smashed* each other ? (To smash : écraser, heurter avec violence).

3) COLLIDE : About how fast were the cars going when they *collided* each other ? (To collide : entrer en collision, s'emboutir).

4) BUMP : About how fast were the cars going when they *bumped* each other ? (To bump : cogner, frapper).

5) CONTACT : About how fast were the cars going when they *contacted* each other ? (To contact : entrer en contact).

Les sujets répondaient en indiquant une vitesse exprimée en miles (nous sommes aux U.S.A). Voici les résultats obtenus (lors d'une réplique de l'expérience) :

HIT	SMASH	COLLIDE	BUMP	CONTACT
22	38	43	47	27
29	40	39	29	24
33	50	32	58	46
50	45	44	34	37
19	48	29	36	31
37	56	44	43	37
33	52	45	25	34
43	47	33	58	18
40	39	48	24	28
34	40	37	31	26

Après avoir identifié les variables dépendante(s) et indépendante(s), vous tirerez les conclusions de cette expérimentation.
Pour vous aider voici quelques statistiques pour chaque groupe :

	$T_{.j}$	$T_{.j}/n_j$	$T_{.j}^2/n_j$	$\Sigma_j x_{ij}^2$
Gr. 1	340	34.0	11560	12338
Gr. 2	455	45.5	20702.5	21043
Gr. 3	394	39.4	15523.6	15894
Gr. 4	385	38.5	14822.5	16241
Gr. 5	308	30.8	9486.4	10060
Total	1882		72095	75576

La Variable Dépendante est évidemment la vitesse exprimée en miles. La Variable Indépendante est le type de verbe utilisé pour poser la question sur la vitesse des voitures.
Manifestement, E. Loftus veut montrer que les "sous-entendus" des verbes sont pris en compte par les sujets dans leur décision sur la vitesse (e.g., les sujets utilisent la signification implicite des verbes comme une source d'information). Le point d'importance dans cette expérience est de remarquer que E. Loftus désire généraliser ses résultats à l'ensemble des verbes signifiant quelque chose comme "entrer en contact". Quoiqu'elle n'ait pas, à proprement parler, sélectionné ses verbes au hasard, elle les juge représentatifs de l'ensemble des verbes de mouvement. Le problème ici est de décider si le facteur expérimental est fixé ou aléatoire. Si l'on admet que les verbes choisis par Loftus représentent un échantillon représentatif, on décidera que le facteur est aléatoire (cf. La polémique initiée par Clark 1973). Si l'on juge que les modalités sont choisies en fait arbitrairement. on décidera que le facteur est fixé, et les conclusions de l'étude se restreignent aux modalités effectivement présentes dans l'expérimentation. Quelle que soit la décision prise, elle sera criticable.
Ici, le distinguo entre facteur fixé et aléatoire peut paraître sans importance car la décision (rejet ou non de l'hypothèse nulle) sera identique dans les deux cas. *Ce ne sera plus le cas dans des plans d'expérience plus complexes.* En fait, l'essentiel de l'argument de Clark (1973) est de montrer qu'une partie des recherches utilisant du matériel linguistique aboutit à des conclusions SCIENTIFIQUES erronées du fait de la confusion entre facteurs fixés et aléatoires (cf. aussi les réponses de Wike et Church 1976). Clark défend l'idée qu'une partie des conclusions de la psychologie du langage est invalide pour avoir cru que des facteurs aléatoires étaient fixes. A cette attaque répond Chastaing (1986) qui démontre méthodologiquement qu'une autre partie de la psychologie du langage est invalide d'avoir cru que des facteurs fixes étaient aléatoires !
Dans le cas présent, le choix entre les deux modèles n'a pas d'influence sur les résultats de l'analyse statistique : on aboutit à des conclusions statistiques identiques (mais pas à des interprétations psychologiques identiques !). L'analyse de variance permet de conclure en tout cas à un effet sur la vitesse estimée, du type le verbe utilisé pour poser la question.
On obtient le tableau d'analyse de variance suivant :

Source	ddl	SC	CM	F_{cal}	$Pr(F > F_{cal})$
Expérimentale	4	1256.52	314.13	4.06 **	.0069
Erreur	45	3481.00	77.36		
Total	49	4737.52			

Ainsi, le type de verbe employé pour interroger les sujets sur la vitesse des véhicules, influence l'estimation qu'ils donnent ($F_{cal}(4, 45) = 4.06,\ p < .05$). On remarque la vitesse élevée induite par *to smash*. Nous pourrions poursuivre cet exemple en essayant d'apprécier les différences entre ces différents verbes les uns par rapport aux autres).

Enoncé 18 *Données Besancon*
On fait subir à 30 élèves d'une école de Besançon une épreuve de "précision perceptive" qui consiste à évaluer un nombre de points sur une diapositive projetée pendant un temps relativement court (une demi-seconde). Les auteurs de cette expérience pensent que la présence d'un témoin peut influencer la performance des sujets dans cette tâche perceptive. Pour vérifier cette idée, les expérimentateurs divisent leur échantillon en trois groupes — chaque enfant étant affecté à un groupe en utilisant une "table de nombres au hasard". Dans le premier groupe (A1) l'expérience est effectuée sans témoin ; dans le second groupe

(A2) l'enfant accomplit sa tâche en compagnie d'un témoin présenté par l'expérimentateur comme un spécialiste ; dans le troisième groupe (A3), le témoin est présenté comme un simple curieux. On répète — pour chaque sujet — vingt-cinq fois l'expérience. Et l'on retient pour chaque sujet la moyenne des écarts absolus (i.e. en ignorant le signe) entre l'estimation fournie et le nombre exact de points.
Les expérimentateurs s'attendent à trouver des différences entre les trois conditions expérimentales ; mais, plus précisément, entre la condition "sans témoin" et la condition "témoin simple curieux" (cette différence leur permettrait de contredire un de leurs collègues qui avançait dans une expérience voisine que le témoin n'agissait que parce que les enfants le jugeait spécialiste). Les auteurs veulent, également vérifier l'existence d'un effet spécifique à la condition "témoin spécialiste".
Questions :
Pourquoi les expérimentateurs décident-ils de prendre l'écart absolu et non pas — par exemple — l'écart relatif. Tout de même, pourquoi retiennent-ils la moyenne des vingt-cinq essais, plutôt qu'un seul essai ?
Quelle est la (les) variable(s) indépendante(s), la (les) variable(s) dépendante(s) ?
Après avoir traduit en termes statistiques les objectifs des expérimentateurs, peut-on penser que ces objectifs sont atteints ? Voici les résultats obtenus :

Condition A1	140	124	118	115	110	110	108	104	102	90
Condition A2	170	164	161	158	156	148	143	140	130	126
Condition A3	136	120	112	104	102	96	92	84	81	75

Eléments de réponses. Calculs intermédiaires :

	A1	*A2*	*A3*	*Totaux*	
n_j	10	10	10	30	
T_j	1121	1496	1002	3619	436572.03
Σx_{ij}^2	127303	225746	103582	456637	
T_j^2/n_j	125664.1	223801.6	100400.4	449866.1	
Inter	13294.07				
Total	20064.97				

Le tableau d'analyse de variance est donné par :

Source	*ddl*	*SC*	*CM*	F_{cal}
Inter-groupes	2	13294.1	6647.03	26.51 **
Intra-groupes	27	6770.9	250.77	
Total	29	20065		

Les trois groupes ne sont donc pas équivalents. La méthode peut être poursuivie en décomposant la variation intra-groupes selon les deux contrastes orthogonaux suggérés par l'énoncé :
$L_1 = \overline{x_1} - \overline{x_3} = 11.9$
$L_2 = 2\overline{x_2} - \overline{x_1} - \overline{x_3} = 86.9$

On obtient alors :
$SC_{contraste1} = 708.05; F = 2.82; Pr(F) = 0.10$
$SC_{contraste2} = 12586.02; F = 50.19; Pr(F) = 1.3 \times 10^{-7}$

L'expérience ne met donc pas de différence en évidence entre les conditons "sans témoins" et "témoin simple curieux" mais par contre, montre un comportement différent dans la condition "témoin spécialiste".

Enoncé 19

Un chercheur a soumis quatre groupes de cinq élèves à un apprentissage de "résolutions de problèmes mathématiques". Chaque groupe apprend avec une méthode pédagogique propre : le premier avec une méthode uniquement verbale, le second avec une méthode écrite, le troisième avec un schéma annoté, le quatrième avec une série de schémas annotés. L'apprentissage dure une heure pour chaque groupe, et le même contenu est présent. Deux jours après l'apprentissage, les sujets sont soumis à un test de raisonnement mathématique. Ce test provient des travaux d'autres chercheurs qui ont étalonné ce test sur une population comparable à celle dont provient l'échantillon d'enfants utilisé ici ; le résultat de ce test est une note (de 0 à 35 : plus la note est élevée, meilleur est le résultat).

Quelle est la Variable Indépendante, la Variable Dépendante ? Comment l'expérimentateur traitera-t-il les résultats de son expérience (souvenez-vous qu'il faut pouvoir répondre à cette question avant de recueillir les résultats !) ?

En outre, l'auteur a mis au point cette expérience pour vérifier certaines hypothèses précises :

1. La méthode verbale diffère-t-elle de l'ensemble des autres méthodes
2. La méthode écrite diffère-t-elle des méthodes avec schémas (un ou plusieurs)' ?
3. Le nombre de schémas a-t-il une influence décelable sur la performance ?

L'auteur peut-il répondre simultanément à ces différentes questions, et quelles seront les réponses ? Interprétez — en vous justifiant — les résultats obtenus et concluez.

Voici les résultats :

GROUPE EXPÉRIMENTAL			
A1	A2	A3	A4
6	14	22	23
13	10	11	19
16	14	19	25
14	19	19	24
14	25	23	25

Elements de réponses. Calculs intermédiaires :

	A1	*A2*	*A3*	*A4*	*Totaux*	
n_j	*5*	*5*	*5*	*5*	*20*	
T_j	*63*	*82*	*94*	*116*	*355*	*6301.25*
Σx_{ij}^2	*853*	*1478*	*1856*	*2716*	*6903*	
T_j^2/n_j	*793.8*	*1344.8*	*1767.2*	*2691.2*	*6597*	
Inter	*295.75*					
Total	*601.75*					

Le tableau d'analyse de variance est donné par :

Source	*ddl*	*SC*	*CM*	F_{cal}	$Pr(F > F_{cal})$
Inter-groupes	*3*	*295.75*	*98.58*	*5.15*	*0.011*
Intra-groupes	*16*	*306*	*19.125*		
Total	*19*	*601.75*			

La méthode peut être poursuivie en décomposant la variation inter-groupes selon les trois contrastes orthogonaux suggérés par l'énoncé :
$L_1 = 3\overline{x_1} - \overline{x_2} - \overline{x_3} - \overline{x_4} = -20.6$
$L_2 = 2\overline{x_2} - \overline{x_3} - \overline{x_4} = -9.2$
$L_3 = \overline{x_3} - \overline{x_4} = -4.4$

On obtient alors :
$SC_{contraste1} = 176.82; F = 9.24; Pr(F) = 0.0067$
$SC_{contraste2} = 70.53; F = 3.69; Pr(F) = 0.07$
$SC_{contraste3} = 48.40; F = 2.53; Pr(F) = 0.13$

Tests non paramétriques sur des groupes indépendants

Enoncé 20
Dans une enquête, on a interrogé 84 hommes et 91 femmes. Les sujets devaient indiquer leur degré d'adhésion à une affirmation, sur une échelle en 5 points. Les résultats sont les suivants :

	Hommes	Femmes
Tout à fait d'accord	10	24
D'accord	15	15
Indifférent	19	21
Opposé	18	17
Tout à fait opposé	22	14

Etudier, à l'aide d'un test de Kolmogorov-Smirnov s'il existe une différence d'opinion entre les hommes et les femmes.
Réponse : Le tableau des fréquences cumulées est donné par :

	Hommes	*Femmes*	*Différence*
Tout à fait d'accord	*0.12*	*0.26*	*0.14*
D'accord	*0.30*	*0.43*	*0.13*
Indifférent	*0.53*	*0.66*	*0.13*
Opposé	*0.74*	*0.85*	*0.10*
Tout à fait opposé	*1.00*	*1.00*	*0.00*

D'où $D = 0.14$*,* $\chi^2 = 3.673$*. Le niveau de significativité vaut ici* $p = 0.15933$*. On n'a donc pas mis en évidence de différence entre les opinions des deux sexes.*

Enoncé 21
In a study of correlates of authoritarian personality structure, one hypothesis was that people high in authoritarianism would show a greater tendency to possess stereotypes about members of various national and ethnic groups than would those low in authoritarianism. This hypothesis was tested with a group of 98 randomly selected college women. Each subject was given 20 photographs and asked to "identify" (by matching) as many or as few photographs as they wished. Since, unknown to the subjects, all photographs were of Mexican nationals – either candidates for the Mexican legislature or winners in a Mexican beauty contest and since the matching list of 20 different national and ethnic groups did not include "Mexican", the number of photographs which any subject "identified" constitutes an index of that subject's tendency to stereotype.
Authoritarianism was measured by the F scale of authoritarianism, and the subjects were grouped as "high" and "low" scorers. High scorers were those who scored at or above the median on the F scale; low scorers were those who scored below the median. The prediction was that these two groups would differ in the number of photographs they identified.

Number of photographs "identified"	Low scorers	High scorers
0-2	11	1
3-5	7	3
6-8	8	6
9-11	3	12
12-14	5	12
15-17	5	14
18-20	5	6

Source : Siegel, S. (1954). Certain determinants and correlates of authoritarianism. Genetic and Psychological Monographs, 49, 187-229.
Comparer les deux groupes à l'aide d'un test unilatéral de Kolmogorov-Smirnov au seuil de 1%.
Indications de réponses : La comparaison des deux fonctions de répartition est donnée dans le tableau suivant :

	0-2	*3-5*	*6-8*	*9-11*	*12-14*	*15-17*	*18-20*
$S_{44}(X)$	*11/44*	*18/44*	*26/44*	*29/44*	*34/44*	*39/44*	*44/44*
$S_{54}(X)$	*1/54*	*4/54*	*10/54*	*22/54*	*34/54*	*48/54*	*54/54*
$S_{44}(X) - S_{54}(X)$	*.232*	*.355*	*.406*	*.252*	*.143*	*.003*	*.0*

D'où $D = 0.406$, $\chi^2 = 4 \times 0.406^2 \frac{44 \times 54}{44+54} = 15.99$. H_1 *est retenue*

Enoncé 22

Dans une étude sur l'agressivité chez les jeunes enfants, un expérimentateur observe des binômes d'enfants dans une situation de jeu contrôlée. Il ne peut étudier que deux enfants par jour, et se demande si un biais pourrait apparaître du fait des discussions entre les enfants observés et ceux qui ne l'ont pas encore été. Si tel est le cas, la distribution des scores, ordonnée selon la date d'observation, ne devrait pas être aléatoire. Pour répondre à cette question, on convertit les scores en une variable dichotomique, selon leur position par rapport à la médiane des valeurs observées. On réalise ensuite un test des suites sur la suite des valeurs observées pour cette variable.
Les données sont les suivantes :

Sujet	Score	Pos/Med
1	31	+
2	23	−
3	36	+
4	43	+
5	51	+
6	44	+
7	12	−
8	26	+
9	43	+
10	75	+
11	2	−
12	3	−

Sujet	Score	Pos/Med
13	15	−
14	18	−
15	78	+
16	24	−
17	13	−
18	27	+
19	86	+
20	61	+
21	13	−
22	7	−
23	6	−
24	8	−

Vérifier que le nombre de "runs" est ici : $u = 10$, et que H_0 peut être retenue pour un test bilatéral au seuil de 5%.
Indications de réponse : Ici, $n_1 = n_2 = 12$. Pour un test bilatéral au seuil de 5%, la table du test des suites donne comme valeurs critiques $u_1 = 7$ et $u_2 = 19$. La règle de décision est donc : H_0 est retenue si $7 < u_{obs} < 19$, ce qui est le cas ici.

Enoncé 23
Un chercheur se demande si l'ordre des hommes et des femmes dans une file d'attente à la caisse d'un théâtre est aléatoire ou non. On a relevé ci-dessous le sexe de 50 clients qui se sont présentés successivement à la caisse :
M F M F M M M F F M F M F M F M M M M F M F M F M M F F F M F M F M F
M M F M M F M M M M F M F M M
Répondre au problème posé à l'aide d'un test des suites.
Indications de réponse : On a ici : $n_1 = 30$ et $n_2 = 20$. Le nombre de runs est $u = 35$. En utilisant l'approximation par une loi normale, on obtient $z_{obs} = 2.83$. On peut donc rejeter l'hypothèse H_0 au seuil de 5% bilatéral.

Enoncé 24
This example is based on a study of gender differences in aggressiveness of four-year-old boys and girls (Siegel, 1956, page 138).
Twelve boys and 12 girls were observed during two 15-minute play sessions; each child's aggressiveness was scored (in terms of frequency and degree) during those sessions and a combined single aggressiveness index was derived for each child.
Ces données se trouvent dans le fichier Aggressn.sta (fichier exemple fourni avec Statistica).

Sexe	Score	Rang
GARCON	86	20
GARCON	69	18
GARCON	72	19
GARCON	65	16.5
GARCON	113	22
GARCON	65	16.5
GARCON	118	23
GARCON	45	12
GARCON	141	24
GARCON	104	21
GARCON	41	11
GARCON	50	13

Sexe	Score	Rang
FILLE	55	14
FILLE	40	10
FILLE	22	7
FILLE	58	15
FILLE	16	4
FILLE	7	1
FILLE	9	2
FILLE	16	5
FILLE	26	8
FILLE	36	9
FILLE	20	6
FILLE	15	3

1) Pourquoi ne semble-t-il pas pertinent d'utiliser un test paramétrique pour comparer ces deux groupes ?
2) Réaliser un test de Wald-Wolfowitz sur ces données. Comparer et vérifier les résultats trouvés avec ceux fournis par Statistica :

```
Test des Suites de Wald-Wolfowitz (Agressn.sta)
Tests significatifs marqués à p <,05000
             Z         niv. p     Z ajusté     niv. p   Nbe de Suites
AGGRESS -3,75681    0,000172    3,548100    0,000388      4
```

3) Réaliser un test de Kolmogorov-Smirnov. Comparer avec les résultats fournis par Statistica :

```
Test de Kolmogorov-Smirnov (Agressn.sta)
Tests significatifs marqués à p <,05000
         Max Nég Differnc   Max Pos Differnc    niv. p
AGGRESS      0,00              0,833333         p < .001
```

4) Réaliser enfin un test de Mann-Whitney.

```
Test U de Mann-Whitney (Agressn.sta)
Tests significatifs marqués à p <,05000
    SommeRgs   SommeRgs   U       Z        niv. p   Z ajusté   niv. p
     GARCON      FILLE
    216,0000    84,00000  6,00 3,810512  0,000139  3,812170   0,000138
```

Enoncé 25

On réalise un test de Mann-Whitney sur deux échantillons de tailles respectives $n_1 = 3$ et $n_2 = 4$. Le protocole des rangs observés sur l'ensemble des 7 observations est le suivant :
Groupe 1 : 1, 2, 4
Groupe 2 : 3, 5, 6, 7
1) Calculer W_1 et W_2. Quelle est la somme des rangs. Comment peut-on la retrouver ?
2) *Passage des statistiques W_1 et W_2 aux statistiques U_1 et U_2.*
Calculer U_1, puis, pour chacun des sujets du groupe 1, compter le nombre de sujets du groupe 2 qui sont classés après lui. Additionner les 3 décomptes obtenus. Que constate-t-on ?
De même, calculer U_2, puis, pour chacun des sujets du groupe 2, compter le nombre de sujets du groupe 1 qui sont classés après lui. Additionner les 4 décomptes obtenus.
3) On veut étudier l'ensemble des protocoles obtenus en affectant au hasard 3 des 7 sujets dans le groupe 1.
a) Il existe 35 protocoles de ce type. Justifier.
b) Pour chacun de ces 35 protocoles, calculer la somme des rangs dans le premier groupe. Représenter la distribution obtenue à l'aide d'un diagramme en bâtons.
c) Quels sont les protocoles pour lesquels on pourrait conclure sur l'hypothèse alternative H_1 au seuil de 5% unilatéral ?
d) Le protocole observé est-il significatif d'une différence entre les deux groupes ?
4) Les échantillons considérés sont évidemment trop petits pour qu'il soit légitime d'utiliser une approximation par une loi normale. Vérifier cependant que les deux formules données dans le cours conduisent à la même valeur de la statistique Z.

Enoncé 26

Un psychologue scolaire veut étudier l'influence du niveau d'étude des mères d'élèves d'une classe de lycéens sur la fréquence de leurs visites auprès de la direction de l'école. Il obtient, pour l'année 1998-1999 :

Niveau d'études	Nombre de visites de la mère
Etudes primaires	4, 3, 0, 7, 1, 2, 0, 3, 5, 1
CEP	2, 4, 1, 6, 3, 0, 2, 5, 1, 2, 1
BEPC	2, 0, 4, 3, 8, 0, 5, 2, 1, 7, 6, 5, 1
Baccalauréat	9, 4, 2, 3
Etudes supérieures	2, 4, 5, 2, 2, 6

1) Montrer que les protocoles de rangs des trois groupes sont donnés par :

```
30.0 25.0 3.0 41.5 9.0 17.5 3.0 25.0 35.0 9.0
17.5 30.0 9.0 39.0 25.0 3.0 17.5 35.0 9.0 17.5 9.0
17.5 3.0 30.0 25.0 43.0 3.0 35.0 17.5 9.0 41.5 39.0 35.0 9.0
44.0 30.0 17.5 25.0
17.5 30.0 35.0 17.5 17.5 39.0
```

2) On demande à un logiciel de traitements statistiques de réaliser un test de Kruskal-Wallis sur ces données. Le résultat produit est le suivant :

```
Kruskal-Wallis rank sum test
data : list(x1, x2, x3, x4, x5) Kruskal-Wallis chi-squared = 2.8532, df = 4,
p-value = 0.5827
```

Refaites les calculs et confirmez les résultats donnés par le logiciel.

3) On réalise le test de Kruskal-Wallis à l'aide de Statistica, qui fournit les résultats suivants :

```
ANOVA de Kruskal-Wallis par Rangs ; Nb Visites (Données Psy-Sco)
Var. indépendante (classement) : Niveau
Test de Kruskal-Wallis : H ( 4, N= 44) =2,853226 p =,582
```

Ces résultats sont-ils en accord avec les précédents ?

4) Le test de la médiane généralisée, réalisé avec Statistica, donne :

```
Test Médiane, Méd. Globale = 2,50000 ; Nb Visites (Données Psy-Sco)
Var. indépendante (classement) : Niveau
Chi-Deux = 1,895105 dl = 4 p = ,7550
```

Vérifiez les résultats et expliquer pourquoi le niveau de significativité du résultat (.75) est nettement plus élevé que celui du test de Kruskal-Wallis (.58).

Enoncé 27

On réalise une enquête sur la satisfaction professionnelle éprouvée par les personnes actives, selon la profession. La satisfaction professionnelle est mesurée par 18 facteurs, sur une échelle de 1 à 5. La somme des évaluations des 18 facteurs est utilisée comme mesure de la satisfaction professionnelle. Une évaluation élevée correspond à un fort degré de satisfaction professionnelle.

1) Pour un échantillon de 10 juristes et un échantillon de 10 analystes informatiques, les données observées sont les suivantes :

Juristes	41	42	42	44	45	48	50	53	64	76
Analystes	38	44	55	60	62	64	66	71	73	74

a) Un statisticien conseille aux auteurs de l'enquête d'utiliser un test non paramétrique pour étudier ces données. Quelles sont les raisons qui l'amènent à faire ce choix ?

b) Etudier, à l'aide d'un test de Wilcoxon-Mann-Whitney, si la satisfaction professionnelle est plus élevée chez les analystes que chez les juristes (test unilatéral au seuil de 5%).

Compte tenu des tailles d'échantillons, on pourra, au choix, utiliser la table du test de Wilcoxon-Mann-Whitney ou l'approximation par une loi normale.

Réponse

a) La variable étudiée est un score numérique calculé à partir de variables ordinales. Bien qu'elle puisse prendre un nombre élevé de valeurs (nombres entiers compris entre 18 et 90), il peut sembler préférable d'utiliser un test ne faisant aucune hypothèse sur la distribution de cette variable dans les populations parentes.

b) On construit tout d'abord le protocole des rangs pour l'ensemble des 20 observations :

Juristes	*2*	*3.5*	*3.5*	*5.5*	*7*	*8*	*9*	*10*	*14.5*	*20*
Analystes	*1*	*5.5*	*11*	*12*	*13*	*14.5*	*16*	*17*	*18*	*19*

La somme des rangs dans le groupe des juristes est 83, celle observée dans le groupe des analystes est 127.

On réalise un test unilatéral en prenant comme hypothèses statistiques :

H_0 : Les scores des juristes et ceux des analystes s'interclassent de manière homogène.

H_1 : La probabilité qu'un score observé chez un juriste soit inférieur à un score observé chez un analyste est supérieure à 50%.

Les deux échantillons sont ici de taille 10. On prend comme valeur observée de la statistique de test $W = 83$. La valeur critique (au seuil de 5%) lue dans la table est : $W_s = 82$.

La règle de décision est donc :

– Si $W \leq 82$, on retient H_1

– Si $W > 82$, on retient H_0.

Par conséquent, on retient H_0.

Variantes de cette solution :

On sait que la statistique U de Mann-Whitney est donnée par :

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - W_2$$

$$U = \min(U_1, U_2)$$

On a donc ici : $U_1 = 155 - 83 = 72$, $U_2 = 155 - 127 = 28$, d'où $U = 28$. La table de la statistique U de Mann-Whitney donne, pour un test unilatéral au seuil de 5%, $U_{crit} = 27$, et la conclusion reste la même.

On peut aussi utiliser l'approximation par une loi normale :

Par exemple : $E^2 = \dfrac{21 \times 20 \times 20}{12 \times 10 \times 10} = 7$; $Z = \dfrac{8.3 - 12.7}{\sqrt{7}} = -1.66.$

Or, pour la loi normale centrée réduite la valeur critique correspondant à un test unilatéral "à gauche" est $Z_c = -1.645$. La conclusion est encore la même.

2) En fait, le tableau précédent ne concernait qu'une partie des données recueillies. L'étude a porté sur 4 professions : juristes, thérapeutes, ébénistes et analystes informatiques. L'ensemble des scores observés est donné par :

Juristes	41	42	42	44	45	48	50	53	64	76
Analystes	38	44	55	60	62	64	66	71	73	74
Thérapeutes	52	55	59	60	62	78	80	86		
Ebéniste	54	59	64	65	69	79	79			

et le protocole des rangs par :

	Juristes	Analystes	Thérapeutes	Ebénistes
	2	1	10	12
	3.5	5.5	13.5	15.5
	3.5	13.5	15.5	22
	5.5	17.5	17.5	24
	7	19.5	19.5	26
	8	22	31	32.5
	9	25	34	32.5
	11	27	35	
	22	28		
	30	29		
$\overline{R}_i$	10.15	18.8	22	23.5

Etudier, à l'aide d'un test de Kruskal-Wallis, si les scores des 4 professions sont significativement différents.

Réponse
Les hypothèses H_0 et H_1 peuvent ici être exprimées par :
H_0 : La probabilité qu'un score provenant de l'une des 4 professions soit supérieur à un score provenant d'une autre profession est de 0.5.
H_1 : Ces probabilités ne sont pas uniformément de 0.5.
La statistique de test K est donnée par :

$$K = \left[\frac{12}{N(N+1)} \sum n_j \overline{R}_j^2\right] - 3(N+1)$$

Elle suit une loi du χ^2 à 3 ddl. La valeur critique, pour un seuil $\alpha = 5\%$ est : $\chi^2_{crit} = 7.815$.
La règle de décision est donc :
– Si $K > 7.815$, on retient l'hypothèse alternative H_1
– Si $K \leq 7.815$, on retient l'hypothèse nulle H_0.
Ici, le calcul donne :

$$K = \frac{12}{35 \times 36}(10 \times 10.15^2 + 10 \times 18.8^2 + 8 \times 22^2 + 7 \times 23.5^2) - 3 \times 36 = 9.16$$

En conclusion, on retient donc l'hypothèse H_1 : les degrés de satisfaction attachés aux 4 professions sont significativement différents.

Tests non paramétriques sur des groupes appareillés

Enoncé 28

Dans le cadre d'une étude sur le tabagisme chez la femme enceinte, on interroge 100 sujets au 3è et au 8è mois de grossesse. On obtient les résultats suivants :

		Fumeur 8è mois	
		oui	non
Fumeur	oui	35	15
3è mois	non	5	45

Le comportement des sujets est-il le même dans les deux conditions ?
Réponse : Le χ^2 de Mac Nemar vaut ici $\chi^2 = 5$, ce qui est significatif d'une différence de comportement au seuil de 5%.

Enoncé 29

14 sujets sont observés dans deux conditions. On obtient 2 différences positives, 10 différences négatives, 2 différences nulles.
Quel est le niveau de significativité obtenu pour un test unilatéral ? pour un test bilatéral ?
Réponses. La statistique de test D_+ suit une loi binomiale de paramètres $N = 12$ et $p = 0.5$.
Calcul du niveau de significativité de $D_{+,obs}$ pour un test unilatéral :

$$P(D_+ = 0) = C_{12}^0\ 0.5^{12} = 0.0002441$$
$$P(D_+ = 1) = C_{12}^1\ 0.5^{12} = 0.0029297$$
$$P(D_+ = 2) = C_{12}^2\ 0.5^{12} = 0.0161133$$

D'où : $p = P(D_+ \leq 2) = 0.019 = 1.9\%$, pour un test unilatéral.
Pour un test bilatéral :
$p = P(D_+ \leq 2) + P(D_+ \geq 10) = 2P(D_+ \leq 2) = 3.8\%$.

Enoncé 30

40 sujets sont observés dans deux conditions. On obtient 10 différences positives, 30 différences négatives, 0 différence nulle.
Le test des signes met-il en évidence une différence de comportement entre les deux conditions ?
Réponse : On a ici : $D_+ = 10$ et $Z = \dfrac{20 + 1 - 40}{\sqrt{40}} = 3.00$
Au seuil de 1% unilatéral, on retient H_1 : les différences négatives sont significativement plus nombreuses que les différences positives.

Enoncé 31

On a testé huit sujets dans deux conditions A_1 et A_2. On obtient le protocole suivant :

Suj.	A_1	A_2
s1	100	105
s2	70	63
s3	40	50
s4	123	98
s5	92	60
s6	120	78
s7	172	119
s8	173	101

Etudier s'il existe une différence significative entre les deux conditions à l'aide d'un test des rangs signés de Wilcoxon.

Réponse. Contruction du protocole des rangs signés :

Suj.	A_1	A_2	d_i	$\|d_i\|$	r_{i+}	r_{i-}
s1	*100*	*105*	*5*	*5*	*1*	
s2	*70*	*63*	*-7*	*7*		*2*
s3	*40*	*50*	*10*	*10*	*3*	
s4	*123*	*98*	*-25*	*25*		*4*
s5	*92*	*60*	*-32*	*32*		*5*
s6	*120*	*78*	*-42*	*42*		*6*
s7	*172*	*119*	*-53*	*53*		*7*
s8	*173*	*101*	*-72*	*72*		*8*
T					*4*	*32*

On trouve $T_+ = 4$, $T_- = 32$ *et donc* $T_m = 4$.
Au seuil de 5% unilatéral, on lit dans la table : $T_{crit} = 5$.
Comme $T_m < T_{crit}$, *on conclut à une différence significative entre les conditions* A_1 *et* A_2 *au seuil de 5% unilatéral.*

Enoncé 32

Nous nous intéressons à l'influence du style d'interview sur les réponses des sujets à une enquête d'opinion. Nous pourrions entraîner un enquêteur à mener trois types différents d'interviews :

– Interview 1 : intérêt, ton amical, enthousiasme,
– Interview 2 : formalisme, réserve, courtoisie,
– Interview 3 : manque d'intérêt, ton abrupt, formalisme pesant.

L'enquêteur visite ensuite trois groupes de 18 foyers, et utilise le style 1 avec un groupe, le style 2 avec le 2è groupe et le style 3 avec le dernier groupe. Nous obtenons ainsi 18 triplets de foyers, comprenant chacun 3 foyers appariés selon des variables pertinentes. Pour chaque triplet, les trois éléments sont affectés au hasard aux trois conditions (styles d'interview). Nous mesurons ensuite l'effet du style d'interview en notant la réponse faite (oui/non) à un item particulier. Les données obtenues sont les suivantes :

Triplet	Style 1	Style 2	Style 3
1	0	0	0
2	1	1	0
3	0	1	0
4	0	0	0
5	1	0	0
6	1	1	0
7	1	1	0
8	0	1	0
9	1	0	0
10	0	0	0
11	1	1	1
12	1	1	1
13	1	1	0
14	1	1	0
15	1	1	0
16	1	1	1
17	1	1	0
18	1	1	0

Etudier à l'aide d'un test Q de Cochran si le style d'interview a une influence sur les réponses obtenues.
Réponse. On obtient ici $G_1 = 13$, $G_2 = 13$, $G_3 = 3$, $G = 29$, $\sum L_i^2 = 63$, *d'où* $Q = 16.7$. *Pour* $ddl = 2$ *et un seuil de 1%, la table donne :* $\chi^2_{crit} = 9.21$. *Les fréquences de réponses positives sont donc différentes selon les styles.*

Enoncé 33
This example is based on a study by Dodd (1979). When processing speech, one actually pays a lot of attention to visual cues as well; specifically, one can understand (encode) spoken words much more readily when the face of the person talking can be seen. In a sense, all people are "lip readers," at least to some extent. Dodd tried to find out whether infants as young as only 10 to 16 weeks old are already aware of the relationship between spoken words and the corresponding movements of the lips (of the speaker). For that purpose, Dodd placed the infants in a room so that they could watch a person behind a window reading normal speech. This speech was either delivered directly into the room (synchronous condition) or it was delayed by 400 milliseconds (asynchronous condition). The dependent variable was the amount of time that the infant watched the face behind the window. No hypotheses were formulated regarding the specific condition that should elicit the most attention (the asynchronous speech could be more interesting because it is novel, or it could draw attention away from the face because the face does not seem to be the source of the speech).

	SYNCHRO	DECALAGE
DC	20.3	50.4
MK	17.0	87.0
VH	6.5	25.1
JM	25.0	28.5
SB	5.4	26.9
MM	29.2	36.6
RH	2.9	1.0
DJ	6.6	43.8
JD	15.8	44.2
ZC	8.3	10.4
CW	34.0	29.9
AF	8.0	27.7

Etudier, à l'aide d'un test des signes, puis d'un test des rangs signés de Wilcoxon, si les données recueillies confirment l'hypothèse du chercheur.

Enoncé 34

Un ergonome désire étudier la forme la plus économique pour un orifice dans lequel des ouvriers doivent faire passer une fiche. Il compare cinq formes d'orifices de moins en moins évasés. Grâce à un appareillage avec cellule photo-électrique, il mesure en millièmes de seconde le temps mis par un ouvrier pour mettre la fiche en position dans l'orifice. Chaque sujet effectue plusieurs essais et l'ergonome note pour chacun le temps médian. Pour 7 sujets, il a obtenu :

Sujet	Taille 1	Taille 2	Taille 3	Taille 4	Taille 5
1	244	417	178	195	452
2	235	307	225	346	613
3	308	290	257	427	438
4	343	305	290	215	534
5	254	263	252	340	469
6	251	291	417	263	445
7	333	414	414	276	441

Etudier, à l'aide d'un test de Friedman, si les médianes correspondant aux 5 formes d'orifices sont égales.

Réponse. Le protocole des rangs par sujet est donné par :

Sujet	*Taille 1*	*Taille 2*	*Taille 3*	*Taille 4*	*Taille 5*
1	*3*	*4*	*1*	*2*	*5*
2	*2*	*3*	*1*	*4*	*5*
3	*3*	*2*	*1*	*4*	*5*
4	*4*	*3*	*2*	*1*	*5*
5	*2*	*3*	*1*	*4*	*5*
6	*1*	*3*	*4*	*2*	*5*
7	*2*	*3.5*	*3.5*	*1*	*5*
R_j	*17*	*21.5*	*13.5*	*18*	*35*
R_j^2	*289*	*462.25*	*182.25*	*324*	*1225*

$$F_r = \frac{12 \times (289 + 462.25 + 182.25 + 324 + 1225)}{7 \times 5 \times 6} - 3 \times 7 \times 6 = 15.86$$

Pour un seuil de 5% et 4 ddl, on a : $\chi^2_{crit} = 9.49$. *On peut donc rejeter l'hypothèse nulle et conclure qu'il y a une influence de la forme de l'orifice sur le temps d'exécution de la tâche.*

Enoncé 35

L'hypnose : dans une expérimentation pratiquée en 1975, Lehman a enregistré le "potentiel cutané" en millivolts chez 8 sujets qui, par ailleurs, étaient interrogés sur la coloration psychique "crainte, joie, tristesse, calme" sous hypnose. Voici le tableau des observations :

	fear	joy	sadness	calmness
1	23.1	22.7	22.5	22.6
2	57.6	53.2	53.7	53.1
3	10.5	9.7	10.8	8.3
4	23.6	19.6	21.1	21.6
5	11.9	13.8	13.7	13.3
6	54.6	47.1	39.2	37
7	21.0	13.6	13.7	14.8
8	20.3	23.6	16.3	14.8

Etudier si l'effet de la coloration psychique sur le potentiel cutané est significatif à l'aide d'un test non paramétrique.

Réponse : le tableau des rangs par sujet s'écrit :

	1	*2*	*4*	*3*
	1	*3*	*2*	*4*
	2	*3*	*1*	*4*
	1	*4*	*3*	*2*
	4	*1*	*2*	*3*
	1	*2*	*3*	*4*
	1	*4*	*3*	*2*
	2	*1*	*3*	*4*
R_j	*13*	*20*	*21*	*26*
R_j^2	*169*	*400*	*441*	*676*

D'où $F_r = \frac{12}{8 \times 4 \times 5}(169 + 400 + 441 + 676) - 3 \times 8 \times 5 = 6.45$

La différence entre les conditions n'est pas significative.

Enoncé 36

Supposons que l'on demande à trois mélomanes d'une revue d'écouter 6 versions différentes d'une symphonie de Beethoven et de les ranger séparément suivant l'organisation des plans sonores (qui ressortissent de l'organisation spatiale des instruments, laquelle varie en général grandement selon le chef d'orchestre). Les trois séries indépendantes de rangs données par les trois mélomanes A, B, C sont exposées dans le tableau suivant :

	a	b	c	d	e	f
A	1	6	3	2	5	4
B	1	5	6	4	2	3
C	6	3	2	5	4	1

Les six versions sont-elles appréciées de la même façon ? Répondre à cette question à l'aide d'un test de Friedman.

Réponse : On a $F = 2.429$. *On n'a pas mis en évidence de différence entre les différentes versions.*